UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/750,011	12/31/2003	Andrew S. Laucius	MS306413.1	9331

27195 7590 05/06/2009

TUROCY & WATSON, LLP
127 Public Square
57th Floor, Key Tower
CLEVELAND, OH 44114

EXAMINER	
SCIACCA, SCOTT M	
ART UNIT	PAPER NUMBER
2446	

NOTIFICATION DATE	DELIVERY MODE
05/06/2009	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

docket1@thepatentattorneys.com
hholmes@thepatentattorneys.com
lpasterchek@thepatentattorneys.com

Office Action Summary	Application No.	Applicant(s)	
	10/750,011	LAUCIUS ET AL.	
	Examiner	Art Unit	
	Scott M. Sciacca	2446	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE *3* MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *16 January 2009*.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-15 and 17-28* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-15 and 17-28* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *31 December 2003* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

This office action is responsive to communications filed on January 16, 2009. Claims 1, 19, 22 and 24-26 have been amended. Claims 1-15 and 17-28 are pending in the application.

Claim Rejections - 35 USC § 103

1. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

2. Claims 1-15, 17, 18, 22, 23 and 26-28 are rejected under 35 U.S.C. 103(a) as being unpatentable over Obata et al. (EP 1120717 A2) in view of Najork et al. (US 6,263,364).

Regarding Claim 1, Obata teaches a system that facilitates incremental web crawls comprising the following components stored in computer memory and executable by a processor:

an indexer that places items into a chunk, wherein the items are the results returned by a web crawl (*"The gatherer process 304 inserts the starting URLs 306 into a transaction log 310. The transaction log 310 identifies those documents that are to be crawled during the current crawl"* – See [0036]); and,

a chunk map that stores at least some of the properties associated with the respective chunk (*"The gatherer process 304 also maintains a history map 308, which contains an ongoing list of all URLs and other historical information that have been accessed during the current Web crawl and previous crawls"* – See [0037]), wherein the properties are at least one of average time between change and average importance of documents in the respective chunk (*"Historical information such as the first access time 422, the last access time 424, the change count 426, and the access count 428 are used in a statistical model for deciding if a document should be accessed during an adaptive incremental crawl"* – See [0046]; *"In accordance with other aspects of the invention, each Web crawl begins with an active probability distribution containing a plurality of probabilities indicative that a document has changed at a given change rate"* – See [0010]), the chunk map employed to facilitate an incremental web re- crawl, wherein the properties of each chunk stored in the chunk map are utilized to determine a re-crawl of all the items in that chunk (*"An adaptive incremental crawl retrieves only documents that may have changed since the previous crawl. The adaptive incremental crawl uses the existing index and history map"* – See [0023]).

Obata does not explicitly teach that items with similar properties are placed into respective chunks and that the properties are shared by all the items in the respective chunk.

However, Najork discloses placing items with similar properties into respective chunks, where the properties are shared by all the items in a

respective chunk (*"the document is assigned to a priority level subqueue based on a predefined set of criteria 282 are satisfied, including but not limited to: … the document's rate of change, based on (a) its modification date and time"* – See Col. 11, lines 48-55).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to modify the system taught by Obata so that items with similar properties are placed in a respective chunk. Motivation for doing so would be to have crawler that can preferentially select a certain set of URL's over others so as to maximize the quality of the information processed by the other applications to which the web crawler passes downloaded documents (See Najork, Col. 3, lines 1-17).

Regarding Claim 2, Obata teaches the items comprising information associated with a Uniform Resource Locator (*"In an actual embodiment of the invention, each entry 510 in the transaction log 310 contains a URL 512 for the document to be processed"* – See [0047]).

Regarding Claim 3, Obata teaches the items comprising at least one of an HTML file, a PDF file, a PS file, a PPT file, an XLS file and a DOC file (*"A Web site may use one or more Web server computers that store and distribute documents in one of a number of formats including the Hyper Text Markup Language (HTML). An HTML document contains text and metadata or commands providing formatting information"* – See [0003]).

Regarding Claim 4, Obata teaches the items receives from a crawler, the crawler responsible for a specific set of Uniform Resource Locators (*"The Web crawl begins by providing the Web crawler with a set of document addresses that act as seeds for the crawl"* – See [0006]).

Regarding Claim 6, Obata teaches a master control process that serves as an interface between a crawler and a re-crawl controller (*"The gatherer process 304 includes one or more worker threads 312 that process a URL"* – See [0037]).

Regarding Claim 7, Obata teaches the master control process maintaining a known chunks table that stores information for components of a system (*"The worker thread 312 passes this list of URLs to the history map 308"* – See [0040]).

Regarding Claim 8, Obata teaches the master control process exposing an interface for communication with a component of the system (See Fig. 3 – Worker Thread 312 exposes an interface for communication with Indexing Engine 208).

Regarding Claim 9, Obata teaches the interface returning a list of chunks the component should have and where to get the chunks (*"The worker thread 312 passes this list of URLs to the history map 308"* – See [0040]).

Regarding Claim 10, Obata teaches the interface returning a list of the chunks that should be actively served by the component (*"The worker thread 312 passes this list of URLs to the history map 308"* – See [0040]).

Regarding Claim 11, Obata teaches the interface returning a range of chunk identifiers to use in building a new chunk by the component (*"The worker thread 312 then passes the list of properties and text to the indexing engine 208"* – See [0041]).

Regarding Claim 12, Obata teaches the interface causing an old chunk to be retired by the system (*"At step 714, the full crawl deletes the old index"* – See [0052]).

Regarding Claim 13, Obata teaches the master control process facilitating movement of chunks from one component to another component (*"The worker thread 312 retrieves a URL from the transaction log 310 and passes the URL to a filter daemon 314"* – See [0037]; *"The filter daemon 314 returns the list of properties and text to the worker thread 312"* – See [0039]; *"The worker thread 312 then passes the list of properties and text to the indexing engine 208"* – See [0041]).

Regarding Claim 14, Obata teaches movement of chunks being based, at least in part, upon at least one of rebalancing index servers after one goes down, re-crawling pages previously crawled, and, restoring a state of a crawler after it has crashed (*"The gatherer process 304 includes one or more worker threads 312 that process a URL until all the URLs in the transaction log 310 have been processed"* – See [0037]; *"A current document at each document address specification listed in the transaction log is retrieved from its Web site and processed"* – See [0021]; *"An adaptive incremental crawl retrieves only documents that may have changed since the previous crawl. The adaptive incremental crawl uses the existing index and history map. The transaction log is selectively seeded with the document address specifications based on a decision whether or not to access a previously retrieved document"* – See [0023]).

Regarding Claim 15, Obata teaches a re-crawl component that employs the chunk map to determine which chunks, if any, to re-crawl at a particular time (*"In an adaptive incremental crawl, once a decision is made to access a previously retrieved document, the document data is retrieved from a Web site if its time stamp is subsequent to the time stamp stored in the Web crawler's history map"* – See [0023]).

Regarding Claim 17, Obata teaches an index chunk that stores information associated with an index of at least some of the items (*"An indexing*

engine 208 is a computer program that maintains an index 210 of documents" – See [0032]).

Regarding Claim 22, Obata teaches a method of performing document re-crawl comprising:

parsing a first chunk for uniform resource locators, wherein the uniform resource locators are stored as a result of one or more web crawls (*"The mechanism of the invention initially performs a first full crawl wherein a transaction log is "seeded" with one or more document address specifications. A current document at each document address specification listed in the transaction log is retrieved from its Web site and processed"* – See [0021]; *"An adaptive incremental crawl retrieves only documents that may have changed since the previous crawl. The adaptive incremental crawl uses the existing index and history map. The transaction log is selectively seeded with the document address specifications based on a decision whether or not to access a previously retrieved document"* – See [0023]);

accessing a chunk map comprising properties associated with the respective chunk (*"The gatherer process 304 also maintains a history map 308, which contains an ongoing list of all URLs and other historical information that have been accessed during the current Web crawl and previous crawls"* – See [0037]]; *"The adaptive incremental crawl uses the existing index and history map"* – See [0023]), wherein the properties are at least one of average time between change or average importance of documents in the respective chunk (*"Historical*

information such as the first access time 422, the last access time 424, the change count 426, and the access count 428 are used in a statistical model for deciding if a document should be accessed during an adaptive incremental crawl" – See [0046]; *"In accordance with other aspects of the invention, each Web crawl begins with an active probability distribution containing a plurality of probabilities indicative that a document has changed at a given change rate"* – See [0010]); and,

periodically determining, based on the properties of each chunk in the chunk map, whether to re-crawl the chunk of data (In the incremental crawl, only documents which are likely to have changed since the last crawl are re-crawled. The likelihood of change is determined based on the properties stored in the history map).

Obata does not explicitly teach that the properties are shared by all the items in the respective chunk.

However, Najork discloses placing items with similar properties into respective chunks such that the properties are shared by all the items in a respective chunk (*"the document is assigned to a priority level subqueue based on a predefined set of criteria 282 are satisfied, including but not limited to: … the document's rate of change, based on (a) its modification date and time"* – See Col. 11, lines 48-55).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to modify the system taught by Obata so that items with

similar properties are placed in a respective chunk for the same reasons as those given with respect to Claim 1.

Regarding Claim 23, Obata teaches the period determination being based, at least in part, upon, at least one of average time between change and average importance of documents comprising a particular chunk (*"Historical information such as the first access time 422, the last access time 424, the change count 426, and the access count 428 are used in a statistical model for deciding if a document should be accessed during an adaptive incremental crawl"* – See [0046]; *"In accordance with other aspects of the invention, each Web crawl begins with an active probability distribution containing a plurality of probabilities indicative that a document has changed at a given change rate"* – See [0010]).

Regarding Claim 26, Obata teaches a system that facilitates increment web crawls comprising the following components stored in computer memory and executable by a processor:

means for placing items into chunks, wherein the items are results returned by a web crawl (*"The gatherer process 304 inserts the starting URLs 306 into a transaction log 310. The transaction log 310 identifies those documents that are to be crawled during the current crawl"* – See [0036]); and,

means for storing at least some of the properties associated with the respective chunk (*"The gatherer process 304 also maintains a history map 308, which contains an ongoing list of all URLs and other historical information that*

have been accessed during the current Web crawl and previous crawls" – See [0037]), and the properties are at least one of average time between change or average importance of documents in the respective chunk (*"Historical information such as the first access time 422, the last access time 424, the change count 426, and the access count 428 are used in a statistical model for deciding if a document should be accessed during an adaptive incremental crawl"* – See [0046]; *"In accordance with other aspects of the invention, each Web crawl begins with an active probability distribution containing a plurality of probabilities indicative that a document has changed at a given change rate"* – See [0010]), and employing the stored properties of each chunk to facilitate an incremental web re-crawl (*"An adaptive incremental crawl retrieves only documents that may have changed since the previous crawl. The adaptive incremental crawl uses the existing index and history map"* – See [0023]).

Obata does not explicitly teach that items with similar properties are placed into respective chunks and that the properties are shared by all the items in the respective chunk.

However, Najork discloses placing items with similar properties into respective chunks, where the properties are shared by all the items in a respective chunk (*"the document is assigned to a priority level subqueue based on a predefined set of criteria 282 are satisfied, including but not limited to: … the document's rate of change, based on (a) its modification date and time"* – See Col. 11, lines 48-55).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to modify the system taught by Obata so that items with similar properties are placed in a respective chunk for the same reasons as those given with respect to Claim 1.

Regarding Claim 27, Obata teaches the items comprising information associated with a Uniform Resource Locator (*"In an actual embodiment of the invention, each entry 510 in the transaction log 310 contains a URL 512 for the document to be processed"* – See [0047]).

Regarding Claim 28, Obata teaches the items comprising at least one of an HTML file, a PDF file, a PS file, a PPT file, an XLS file and a DOC file (*"A Web site may use one or more Web server computers that store and distribute documents in one of a number of formats including the Hyper Text Markup Language (HTML). An HTML document contains text and metadata or commands providing formatting information"* – See [0003]).

3. Claims 5 and 18 are rejected under 35 U.S.C. 103(a) as being unpatentable over Obata et al. (EP 1120717 A2) in view of Najork et al. (US 6,263,364) and further in view of Dean et al. (US 7,305,610).

Regarding Claim 5, Obata does not explicitly teach a master control process that can modify the chunk map to facilitate load balancing amongst a

plurality of crawlers. However, Dean does teach facilitating load balancing among a plurality of crawlers (*"FIG. 7 shows a flow chart of a process of adjusting stall times"* – See Col. 7, lines 34-35; *"Once the actual retrieval time is determined, the stall time for the selected host can be adjusted according to the retrieval time at a step 503"* – See Col. 7, lines 40-43; *"each computer system 1 can be executing one or more web crawler that traverses hyperlinked documents and saves information regarding the traversed hyperlinked documents on the computer system"* – See Col. 3, lines 60-63).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Obata to include load balancing among a plurality of crawlers. Motivation for doing so would be to allow for more efficient crawling of documents without placing unnecessary burdens on hosts (See Dean, Col. 1, lines 49-52)

Regarding Claim 18, Obata does not explicitly teach a rank chunk that stores a static rank associated with an index chunk. However, Dean teaches a rank chunk that stores a static rank associated with an index chunk (*"there may be PageRank processes 219 that retrieve links from links files 217 and provides the links with a priority or rank"* – See Col. 5, lines 65-67).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Obata to include a rank chunk. Motivation for doing so would be to assign priorities to links that are to be crawled (See Dean, Col. 5, lines 48-54).

4. Claims 19-21 are rejected under 35 U.S.C. 103(a) as being unpatentable over Dean et al. (US 7,305,610) in view of Evans et al. (US 2004/0030683).

Regarding Claim 19, Dean teaches a method of performing document re-crawl comprising:

parsing a first chunk for uniform resource locators (*"The present invention provides innovative techniques for crawling of hyperlinked documents"* – See Col. 1, lines 41-42; *"At a step 401, links to hyperlinked documents are received. The links are to hyperlinked documents that are to be crawled. The links to hyperlinked documents are grouped by host at a step 403"* – See Col. 7, lines 21-24; *"a single link (e.g., uniform resource locator or URL)"* – See Col. 1, lines 46-47; *"A link (e.g., URL) server 201 determines which links should be crawled next. Link server 201 maintains a pool of uncrawled links and groups the links by the host on which each link resides"* – See Col. 4, lines 13-16; Links to hyperlinked documents having a common host are grouped together (in a chunk). A link server determines which links should be crawled next), wherein a chunk map that stores properties associated with the respective chunk stored in a chunk table is employed to determine the first chunk (*"In order to accomplish rate limiting of hosts, each host has an associated stall time, which is the earliest time at which another link from this host should be crawled or released to a crawler"* – See Col. 6, lines 47-50; The stall times are a property associated with a respective host. The stall time is used to determine which host, and thus, which group of links

(chunk) should be crawled next), wherein the stored properties are shared by all the items in the respective chunk (Since all the links (items) belonging to a particular host are grouped together and each host has an associated stall time (property), the stall time is shared by all links in the group), and

re-crawling the uniform resource locators (*"Once the host to be crawled next is selected, a hyperlinked document from the selected host is crawled at a step 407"* – See Col. 7, lines 27-29).

Dean does not explicitly teach forming a second chunk separate from the first chunk, based at least in part, upon the re-crawled uniform resource locators.

However, Dean does mention that web pages may include URLs which point to another web page on another host (*"In a wide area network such as the Internet, some of the computer systems are servers (or hosts)"* – See Col. 3, lines 39-40; *"The web pages typically include links in the form of uniform resource locators (URLs) that are a link to another web page, whether it is on the same server or a different one"* – See Col. 3, lines 44-47). Thus, in the process of crawling a web page of a first host, a crawler may find a link to a web page hosted by a second host.

Evans teaches a crawler that, upon encountering a new web site, will perform an exhaustive search of the site (*"the first time a web site (or any location of content, such as a file directory) is encountered, an exhaustive search is conducted"* – See [0023]).

Combining these features would yield a crawler that, upon encountering a link to a web page on a second host, would also perform a crawl of the web page

on the newly discovered second host. This would also result in the formation of a second chunk, since the links encountered on the new host would be grouped together in a new chunk.

It would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Dean's method of performing web crawls to include upon encountering a link to a web page on a second host, performing a crawl of the web page on the newly discovered second host. According to Evans, new content is continuously appearing on the web (*"The volume and variety of informational content available on the web is likely to continue to increase at a rather substantial pace"* – See [0002]). It is the job of a crawler to index the content of the web so that search engines may inspect the index and return search results to a user based on a search query (See [0004]). Thus, performing a crawl of a web page on a newly discovered host (and subsequently indexing the contents of the web page) would ensure that content which is new to the web will be available to search engines.

Regarding Claim 20, Dean teaches moving the first chunk (*"When a crawler needs one or more links to crawl, the crawler requests one or more links from link server 201"* – See Col. 4, lines 26-27).

Regarding Claim 21, Dean teaches one or more computer readable media having stored thereon computer executable instructions for carrying out the method of claim 19 (*"FIG. 1 illustrates an example of a computer system that can*

be used to execute the software of an embodiment of the invention" – See Col. 2, lines 65-67; *"Although CD-ROM 15 is shown as an exemplary computer readable storage medium, other computer readable storage media including floppy disk, tape, flash memory, system memory, and hard drive can be utilized. Additionally, a data signal embodied in a carrier wave (e.g., in a network including the Internet) can be the computer readable storage medium"* – See Col. 3, lines 8-14).

5. Claims 24 and 25 are rejected under 35 U.S.C. 103(a) as being unpatentable over Obata et al. (EP 1120717 A2) in view of Najork et al. (US 6,263,364) and Dingsor et al. (US 7,058,727).

Regarding Claim 24, Obata teaches a computer-readable medium encoded with data structure transmitted between two or more computer components that facilitates document re-crawl, the data structure comprising:

a chunk header that includes metadata associated with the data structure (*"After retrieving a document, the filter daemon 314 parses the document and returns a list of text and properties"* – See [0038]; *"The information may be text to be displayed, "metadata" that describes the formatting of the text, hyperlinks, or other information"* – See [0038]), the chunk comprising document files associated with one or more uniform resource locators (*"A history map is maintained by the Web crawler that references a number of documents that were accessed during previous Web crawls"* – See [0010]; *"The document URL, its hash value, its time*

stamp, its crawl number modified and other historical information (discussed below) are stored in a persistent history map" – See [0021]); and

document files that include content found on the Internet (*"The methods and systems of the present invention optimize a Web crawler's use of computer resources when performing adaptive incremental Web crawls to maintain the synchronization between local data copied from a document when it was previously retrieved and current data contained in the document which may have been changed since the document was last retrieved"* – See [0009]),

wherein the average of the at least one of the properties of all the document files determines if the document should be re-crawled (*"Historical information such as the first access time 422, the last access time 424, the change count 426, and the access count 428 are used in a statistical model for deciding if a document should be accessed during an adaptive incremental crawl"* – See [0046]; *"In accordance with other aspects of the invention, each Web crawl begins with an active probability distribution containing a plurality of probabilities indicative that a document has changed at a given change rate"* – See [0010]).

Obata does not explicitly teach that the metadata is shared by all the items in a chunk.

However, Najork discloses placing items with similar metadata into respective chunks (*"the document is assigned to a priority level subqueue based on a predefined set of criteria 282 are satisfied, including but not limited to: ... the*

document's rate of change, based on (a) its modification date and time" – See Col. 11, lines 48-55).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to modify the system taught by Obata so that items with similar properties are placed in a respective chunk for the same reasons as those given with respect to Claim 1.

Obata does not explicitly teach the data structure having an offset section that provides offset information associated with the document files.

However, Dingsor teaches a data structure having an offset section that provides information associated with the data (*"Other fields in the IP header, like total length and fragment offset, are used to breakup network datagrams into packets at the source computer and reassemble them at the destination computer"* – See Col. 6, lines 44-47).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to assemble the metadata and document files taught by Obata along with the offset information taught by Dingsor into a data structure. In order for two computers to communicate it is necessary to place the information in a data packet, such as the IP packet taught by Dingsor. Motivation for doing so would be to allow the information to be transmitted from one computer to another.

Regarding Claim 25, Obata teaches at least one of the document files comprising at least one of an HTML file, a PDF file, a PS file, a PPT file, an XLS

file and a DOC file (*"A Web site may use one or more Web server computers that store and distribute documents in one of a number of formats including the Hyper Text Markup Language (HTML). An HTML document contains text and metadata or commands providing formatting information"* – See [0003]).

Response to Arguments

6. Applicant's arguments with respect to Claims 1-15, 17, 18 and 22-28 have been considered but are moot in view of the new grounds of rejection.

7. Applicant's arguments with respect to Claims 19-21 have been fully considered but they are not persuasive.

On page 10 of the remarks, Applicant argues "Dean et al. is silent regarding performing a re-crawl, and thus does not disclose re-crawling the uniform resource locators."

At Col. 5, lines 23-29, Dean discloses assigning states to each link that has been crawled. If a link is crawled and found to be unreachable, then the crawling is attempted again at a later time. Thus, a link is re-crawled.

On page 10 of the remarks, Applicant argues "At page 16 of the Office Action, the Examiner contends that when performing a crawl, if the crawler

encounters a new web site, it will create a new group. However, this new group is not formed based upon the re-crawled uniform resource locators."

As shown above, in Col. 5, lines 23-29, Dean discloses performing re-crawling on URLs. As shown on page 16 of the Office Action dated 10/16/2008, Dean discloses that links to a different host may be encountered while a URL is being crawled. Links to a different host may be encountered when crawling a URL regardless of whether it is being crawled a first time or a second time after a first failed attempt. Thus, the proposed combination of Dean and Evans with respect to "forming a second chunk separate from the first chunk" meets the claimed invention since this may occur based not only on an initial crawl of a URL, but also when a URL is being re-crawled.

On page 11 of the remarks, Applicant argues "Claim 20 recites *determining whether any chunks are to be retired; moving the first chunk; and, destroying the first chunk*. Dean et al. and Evans do not disclose such novel features."

Applicant's arguments fail to comply with 37 CFR 1.111(b) because they amount to a general allegation that the claims define a patentable invention without specifically pointing out how the language of the claims patentably distinguishes them from the references.

Conclusion

8. Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Scott M. Sciacca whose telephone number is (571) 270-1919. The examiner can normally be reached on Monday thru Friday, 7:30 A.M. - 5:00 P.M. EST.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Jeff Pwu can be reached on (571) 272-6798. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/Scott M. Sciacca/
Examiner, Art Unit 2446

/Jeffrey Pwu/
Supervisory Patent Examiner, Art Unit 2446